Discrete Optimization 9 (2012) 205–208

Contents lists available at SciVerse ScienceDirect

Discrete Optimization

journal homepage: www.elsevier.com/locate/disopt

Note

Analysis of an approximate greedy algorithm for the maximum edge clique partitioning problem

Abraham P. Punnen *, Ruonan Zhang

Department of Mathematics, Simon Fraser University Surrey, Central City, 250-13450 102nd AV, Surrey, British Columbia,V3T 0A3, Canada

ARTICLE INFO

Article history:
Received 24 March 2011
Received in revised form 4 February 2012
Accepted 10 May 2012
Available online 29 May 2012

Keywords:
Maximum clique
Clique partition
Approximation algorithm

ABSTRACT

In this note, we show that if the maximum clique problem can be solved by a polynomial time δ-approximation algorithm, then the maximum edge clique partitioning problem (Max-ECP) can be solved by a polynomial time $\frac{2(p\delta-1)}{p-1}$-approximation algorithm for any fixed integer $p \geq 2$. This improves the best known bound on the performance ratio of an approximation algorithm for Max-ECP problem and also corrects an error in an earlier work on the topic.

1. Introduction

Let $G = (V, E)$ be a graph on n nodes. Then the maximum edge clique partitioning problem (Max-ECP) is to find a partition $V_1, V_2, \dots, V_m$ of V such that the subgraph $G_i = (V_i, E_i)$ of G induced by V_i is a clique for $1 \leq i \leq m$ and $\sum_{i=1}^{m} |E_i|$ is maximized.

Max-ECP was considered by Dessmark et al. [1] and proposed a polynomial time approximation algorithm with performance ratio n. They also gave another approximation algorithm with worst case performance ratio 2. However, this algorithm solves a sequence of maximum clique problems and hence polynomiality is guaranteed only for instances where the associated maximum clique problems can be solved in polynomial time. Recently, Zhang et al. [2] have studied Max-ECP in the context of the minimum spanning tree problem with conflict pairs. They noted that if the maximum clique problem on G and its subgraphs can be solved by a polynomial time δ-approximation algorithm, then the Max-ECP can be solved by a polynomial time $(3\delta - 1)$-approximation algorithm. While this result is correct, the proof given in [2] had an error. Following the analysis given in [2] and correcting an algebraic error, a performance ratio bound of $4\delta - 2$ follows instead of $3\delta - 1$ as indicated in [2]. Modifying the computational scheme of [2], we give a 2-phase approximate greedy algorithm to solve Max-ECP which yields the $3\delta - 1$ bound as a special case. In each iteration, the algorithm extracts a reasonably large size clique in phase 1. The transition from phase 1 to phase 2 is controlled by a parameter p which guarantees that the cliques extracted in phase 1 are of size (number of nodes) at least p. An upper bound on the performance ratio of the 2-phase approximate greedy algorithm is $\frac{2(\alpha\delta-1)}{\alpha-1}$, where α is the average size of the cliques extracted in phase 1 of the algorithm and δ is the performance ratio of the approximation algorithm used to solve the maximum clique problem. A data independent bound on the performance ratio is given by $\frac{2(p\delta-1)}{p-1}$ where $p \geq 2$ is a fixed integer. When $p = 2$, the 2-phase approximate greedy algorithm reduces to the approximate greedy algorithm of [2] and the performance bound established becomes $(4\delta - 2)$. When $p = 3$, the performance ratio coincides with the $(3\delta - 1)$ bound claimed in [2] but the algorithm is slightly different from what is given in [2] to achieve this bound. As α (or p) increases our bound on the performance ratio gets better but the

* Corresponding author.
E-mail addresses: apunnen@sfu.ca (A.P. Punnen), rza1@sfu.ca (R. Zhang).

1572-5286/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.disopt.2012.05.002

complexity of the algorithm also grows with p. Since the maximum clique problem can be approximated within a factor of $\frac{Kn(\log\log n)^2}{(\log n)^3}$ for an appropriate constant K [3], Max-ECP can be approximated within a factor of $\frac{2}{p-1}(p\frac{Kn(\log\log n)^2}{(\log n)^3}-1)$ for any fixed integer $p \geq 2$. This improves the best known performance bound for a polynomial time approximation algorithm for Max-ECP. It may be noted that the asymptotic performance ratio in this case is $O\left(\frac{n(\log\log n)^2}{(\log n)^3}\right)$, which improves the bound of [1], while have the same asymptotic bound as given in [2] after the correction discussed above is applied. We also observe that Max-ECP is NP-hard on graphs with clique number no more than three but polynomially solvable on graphs with clique number no more than two.

Throughout this note, we use the notations $V(G)$ and $E(G)$ to represent respectively the node set and the edge set of a graph G.

2. The 2-phase approximate greedy algorithm

This is a generalization of the approximate greedy algorithm given in [2]. In phase 1, we iteratively extract a clique D^k using an approximate algorithm for the maximum clique problem such that $|V(D^k)|$ is no less than p for a given integer $p \geq 2$. The process is continued until no such cliques can be identified and the algorithm switches to phase 2. In phase 2, since the remaining graph contains only cliques with size at most $p-1$, an exact maximum clique can be found in polynomial time for fixed p. We keep extracting maximum cliques until the edge set of the remaining graph becomes empty. A formal description of the algorithm is given below. Here Approx-Max-Clique(G) is a procedure which accepts G as input and outputs an δ-optimal solution for the maximum clique problem on G.

Algorithm 1: The 2-phase Approximate Greedy Algorithm

```
Input: G = (V, E);
Phase 1:
H^1 = Ø, H^2 = Ø, G^1 = G, k = 0;
while G^{k+1} contains cliques of size ≥ p do
   k = k + 1;
   D^k = Approx-Max-Clique(G^k);
   If |V(D^k)| < p then choose a clique of size p in G^k and designate it D^k;
   H^1 = H^1 ∪ D^k;
   G^{k+1} = G^k \ V(D^k);
end while;
Phase 2:
k = k + 1;
while E(G^k) ≠ Ø do
   D̄^k = Max-Clique(G^k);    /* Max-Clique (G) computes a maximum clique of G */
   H^2 = H^2 ∪ D̄^k;
   G^{k+1} = G^k \ V(D̄^k);
   k = k + 1;
end while;
H = H^1 ∪ H^2;
Output: H ∪ G^k.
```

Note that the maximum clique extraction in phase 2 can be done in $O(n^p)$ time by complete enumeration and hence it is polynomial for fixed p. Also the condition of the while loop in phase 1 can be verified in $O(n^p)$ time by complete enumeration. Thus the algorithm is polynomially bounded whenever Approx-Max-Clique(G) in phase 1 is polynomially bounded. In the 2-phase approximate greedy algorithm, if we let $p = 2$, then phase 2 is redundant and the algorithm reduces to the approximate greedy algorithm introduced in [2].

Lemma 1. *Let $a_1, a_2, \ldots, a_n$ be n non-negative real numbers with mean $\bar{a} > 1$. Then $\sum_{i=1}^n a_i \leq \frac{1}{\bar{a}-1}\sum_{i=1}^n a_i(a_i-1)$.*

Proof. By the sum of squares inequality, we have $\frac{1}{n}(\sum_{i=1}^n a_i)^2 \leq (\sum_{i=1}^n a_i^2)$. Subtracting $\sum_{i=1}^n a_i$ from both sides and simplifying, we get $(\sum_{i=1}^n a_i)(\frac{1}{n}\sum_{i=1}^n a_i - 1) \leq \sum_{i=1}^n (a_i^2 - a_i)$. Note that $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$ and hence $(\bar{a}-1)\sum_{i=1}^n a_i \leq \sum_{i=1}^n a_i(a_i-1)$. Since $\bar{a} > 1$ the result follows. □

Theorem 2. *If Approx-Max-Clique(G) computes an δ-optimal solution to the maximum clique problem on G, then the 2-phase approximate greedy algorithm gives a $\frac{2(\alpha\delta-1)}{\alpha-1}$-optimal solution to Max-ECP on G, where α is the average size of cliques extracted in phase 1.*

Proof. Let $Q^1 = (Q_1^1, Q_2^1, \ldots, Q_t^1)$ be an optimal solution to Max-ECP on G. Also, let r be the number of iterations in phase 1 and s be the total number of iterations in the algorithm. In each iteration k of phase 1, let $|V(D^k)| = d^k$ and ρ^k be the size of a maximum clique of G^k. Since D^k is an δ-optimal solution to the maximum clique problem on G^k, $d^k \le \rho^k \le \delta d^k$ for $k \le r$. Let $Q^k = (Q_1^k, Q_2^k, \ldots, Q_t^k)$ for $1 \le k \le s$ and $Q_i^{k+1} = Q_i^k - V(D^k)$ for $1 \le k \le r$. Note that for $k \le r$, $|E(Q_i^k)| - |E(Q_i^{k+1})| \le |V(Q_i^k) \cap V(D^k)|(\rho^k - 1)$ if $V(Q_i^k) \cap V(D^k) \neq \emptyset$ and $|E(Q_i^k)| - |E(Q_i^{k+1})| = 0$ if $V(Q_i^k) \cap V(D^k) = \emptyset$. Also, $\sum_{i=1}^t |V(Q_i^k) \cap V(D^k)| = d^k$ for $k \le r$. Thus

$$|E(Q^k)| - |E(Q^{k+1})| = \sum_{i=1}^{t} (|E(Q_i^k)| - |E(Q_i^{k+1})|) \le d^k(\rho^k - 1) \le d^k(\delta d^k - 1), \quad \text{for } k \le r. \tag{1}$$

Adding (1) for $k = 1$ to r we get

$$\sum_{k=1}^{r} \left(|E(Q^k)| - |E(Q^{k+1})|\right) \le \sum_{k=1}^{r} d^k(\delta d^k - 1). \tag{2}$$

In phase 2, an exact maximum clique $\bar{D}^k$ is extracted in each iteration k, For $r < k \le s - 1$, let $Q_i^{k+1} = Q_i^k - V(\bar{D}^k)$ and now, $\sum_{i=1}^t |V(Q_i^k) \cap V(\bar{D}^k)| = \rho^k$ for $r < k \le s$. Thus

$$|E(Q^k)| - |E(Q^{k+1})| \le \rho^k(\rho^k - 1), \quad \text{for } k = r + 1, \ldots, s. \tag{3}$$

Adding (3) from $k = r + 1$ to s, we have

$$\sum_{k=r+1}^{s} \left(|E(Q^k)| - |E(Q^{k+1})|\right) \le \sum_{k=r+1}^{s} \rho^k(\rho^k - 1). \tag{4}$$

From (2) and (4), we get

$$\begin{aligned}
\sum_{k=1}^{s}(|E(Q^k)| - |E(Q^{k+1})|) &\le \sum_{k=1}^{r} d^k(\delta d^k - 1) + \sum_{k=r+1}^{s} \rho^k(\rho^k - 1) \\
&= \delta \sum_{k=1}^{r} d^k \left(d^k - 1 + 1 - \frac{1}{\delta}\right) + \sum_{k=r+1}^{s} \rho^k(\rho^k - 1) \\
&= \delta \sum_{k=1}^{r} d^k(d^k - 1) + (\delta - 1) \sum_{k=1}^{r} d^k + \sum_{k=r+1}^{s} \rho^k(\rho^k - 1).
\end{aligned}$$

By definition, $\alpha = \frac{1}{r} \sum_{k=1}^r d^k$ and using Lemma 1 we have

$$\begin{aligned}
\sum_{k=1}^{s}(|E(Q^k)| - |E(Q^{k+1})|) &\le \delta \sum_{k=1}^{r} d^k(d^k - 1) + \frac{\delta - 1}{\alpha - 1} \sum_{k=1}^{r} d^k(d^k - 1) + \sum_{k=r+1}^{s} \rho^k(\rho^k - 1) \\
&= \frac{\alpha\delta - 1}{\alpha - 1} \sum_{k=1}^{r} d^k(d^k - 1) + \sum_{k=r+1}^{s} \rho^k(\rho^k - 1) \\
&= \frac{2(\alpha\delta - 1)}{\alpha - 1} |E(H^1)| + 2|E(H^2)| \\
&\le \frac{2(\alpha\delta - 1)}{\alpha - 1} (|E(H^1)| + |E(H^2)|) \\
&= \frac{2(\alpha\delta - 1)}{\alpha - 1} |E(H)|.
\end{aligned}$$

On the left hand side,

$$\sum_{k=1}^{s}(|E(Q^k)| - |E(Q^{k+1})|) = |E(Q^1)| - |E(Q^{s+1})| = |E(Q^1)| \quad \text{since } E(Q^{s+1}) = \emptyset.$$

Therefore, $|E(Q^1)| \le \frac{2(\alpha\delta-1)}{\alpha-1} |E(H)|$, where $|E(Q^1)|$ is the optimal objective function value of the Max-ECP on G and $|E(H)|$ is the objective function value of the approximate solution returned by Algorithm 1. $\square$

The bound established above contains α which is data dependent. Let us now consider a data independent bound.

Lemma 3. *For $a \ge b > 1$ and $\delta \ge 1$, $\frac{a\delta-1}{a-1} \le \frac{b\delta-1}{b-1}$.*

Proof. If $\delta = 1$, the proof is trivial. Assume $\delta > 1$. Since $a \ge b > 1$, we have $1 - \frac{1}{a} \ge 1 - \frac{1}{b}$ and hence $\frac{a}{a-1} \le \frac{b}{b-1}$. Also $\delta > 1$. Thus $\frac{a}{a-1}(\delta - 1) + 1 \le \frac{b}{b-1}(\delta - 1) + 1$. Simplifying this inequality yields the required result. $\square$

Note that the average size of the cliques extracted in phase 1 is at least p. Thus $\alpha \geq p \geq 2$. From Lemma 3 and Theorem 2, we have that the performance ratio of the 2-phase approximate greedy algorithm is bounded by $\frac{2(p\delta-1)}{p-1}$. Therefore, the 2-phase approximate greedy algorithm has a performance ratio bound of $4\delta - 2$ when $p = 2$ and $3\delta - 1$ when $p = 3$. As the value of α increases, the performance ratio bound of Algorithm 1 approaches 2δ.

Note that the maximum clique problem on a graph can be solved in polynomial time by complete enumeration if its clique number is fixed. However, it appears that such an approach does not extend to Max-ECP.

Theorem 4. *Max-ECP is NP-hard on graphs with clique number no more than* 3 *but solvable in polynomial time on graphs with clique number no more than* 2*.*

Proof. We reduce the 3-dimensional matching problem (3DM) to Max-ECP on a graph with clique number no more than 3. Let X, Y, Z be disjoint sets with $|X| = |Y| = |Z| = n$ and $S \subseteq X \times Y \times Z$. Then the 3DM is to find a subset M of S such that for any two distinct triples (x_1, y_1, z_1) and (x_2, y_2, z_2) in M we have $x_1 \neq x_2, \ y_1 \neq y_2$, and $z_1 \neq z_2$ and $|M| = n$. Given an instance X, Y, Z, S of 3DM, construct a tripartite graph G with partite sets X, Y, and Z. For each triple $(x, y, z) \in S$, there are edges $(x, y), \ (y, z)$ and (x, z) in G. Clearly, the clique number of G is no more than 3. Note that $3n$ is an upper bound on the optimal objective function value of Max-ECP on G and a solution to max-ECP achieves this value precisely when the given instance of 3DM has a 3-dimensional matching M with $|M| = n$. Since 3DM is NP-hard [4] Max-ECP is also NP-hard.

For graphs with clique number no more than two, it can be verified that a maximum cardinality matching solves Max-ECP. $\square$

Acknowledgment

This work was supported by an NSERC discovery grant awarded to Abraham P. Punnen.

References

[1] A. Dessmark, J. Jansson, A. Lingas, E.M. Lundell, M. Persson, On the approximability of maximum and minimum edge clique partition problems, International Journal of Foundations of Computer Science 18 (2007) 217–226.
[2] R. Zhang, S. Kabadi, A. Punnen, The minimum spanning tree problem with conflict constraints and its variations, Discrete Optimization 8 (2011) 191–205.
[3] U. Feige, Approximating maximum clique by removing subgraphs, SIAM Journal on Discrete Mathematics 18 (2005) 219–225.
[4] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman & Co., 1979.